Human β-Casein

AMINO ACID SEQUENCE AND IDENTIFICATION OF PHOSPHORYLATION SITES*

Rae Greenberg, Merton L. Groves, and Harold J. Dower

From the Eastern Regional Research Center, United States Department of Agriculture, Philadelphia, Pennsylvania 19118

The primary structure of human β-casein has been determined by automated Edman degradation of the intact protein and of peptides derived therefrom by hydrolysis with trypsin and by chemical cleavage with cyanogen bromide. For each form of this multiphosphorylated protein (0–5 P/molecule), phosphorylated sites at specific seryl and threonyl residues have been identified. These are located near the amino terminus, within the first 10 residues of this 212-amino acid molecule. Sequence comparison of human β-casein with the bovine and ovine proteins reveals 50% identity and a 10-residue shifted alignment relationship. Locations of prolyl and charged residues are generally conserved for the three homologues. The sequence data indicate the existence of genetic polymorphism involving uncharged residues in human β-casein.

Mature human milk, the food widely recommended for the full term infant by the nutritional and pediatric communities (1, 2), contains 0.9% protein (3), of which the casein fraction comprises 30–35% (4). The major component of human casein has an apparent electrophoretic and compositional similarity to β-casein from the well characterized bovine milk system (5, 6) which contains 3.3% protein, 80% casein, 38% of which is β-casein. Detailed analyses of the human casein fraction including structure and physicochemical properties of β-casein are lacking. At the molecular level, human and bovine β-casein differ significantly, the most obvious difference being the contrast in phosphorylation state. Human β-casein occurs in multiphosphorylated forms having 0–5 phosphate groups/molecule (5, 6), while the bovine homologue is usually found as a fully phosphorylated molecule. Since milk caseins represent the prime source of calcium and phosphorus for the neonate, differences in the concentration and availability of these elements may have implications in the feeding of banked human milk to preterm or at-risk infants (7–10).

To characterize more fully human β-casein at the molecular level, this report presents the complete amino acid sequence of the 212-residue protein and the identification of phosphorylation sites in the individual forms of this multiphosphorylated protein. A comparison of the primary structures of human, bovine (11), and ovine (12) β-caseins reveals both sequence homology and an interesting shifted alignment relationship. The sequence data also indicate the existence of genetic polymorphism in human β-casein. Brief accounts of portions of this work have appeared (13, 14).

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

EXPERIMENTAL PROCEDURES[1]

RESULTS

The β-casein fraction of human milk, isolated by Groves and Gordon (5), contains six electrophoretic bands as shown in Fig. 1. This complex was fractionated and after purification by DEAE-cellulose column chromatography, each of the individual components was found to be identical in amino acid composition (5). These proteins differed only in phosphorus content, values ranging from 0 to 5 mol/mol as indicated in Fig. 1. Previous results from this laboratory (14) had shown that the nonphosphorylated form contained a sequence cluster identical to the known phosphorylated regions of bovine α_{s1}- and β-caseins (15). This suggested that the phosphorylated residues were located near the amino terminus of the molecule. Sequence examinations of the individual 2-P and 4-P β-casein forms confirmed this (14). It was thus determined that each of the multiphosphorylated forms contained phosphate groups on specific seryl or threonyl residues and were not mixtures of species having a certain number of phosphate groups randomly distributed. Using the 0-P amino-terminal sequence as a guide, phosphorus was located in each of the other forms as described under "Experimental Procedures." In all five cases phosphate groups were quantitatively located at specific sites within the first 10 residues (Fig. 2). A complete explanation of the 1-P case is presented in the discussion.

The amino acid sequence of human β-casein and the strategy employed in its elucidation are shown in Fig. 3. The basis for this sequence was provided by the arrangement of cyanogen bromide-cleaved and trifluoroacetyl-blocked trypsin cleaved peptides reported earlier (13) and described in Fig. 1S and Table IS. Peptide CB-1 was digested with trypsin yielding two fragments, CB-1-T1 and CB-1-T2 (Table IS). The first peptide corresponded to residues 1-20 and the second, a larger fragment, to the remainder of CB-1, residues 24–84. The tripeptide 21–23 (Val-Glu-Lys) was not recovered from this digest. Although CB-1 contains five lysyl residues, primary cleavage occurred at the two Lys-Val bonds.

The amino-terminal third of the molecule is encompassed by tryptic fragments T-1A and T-1B, originally thought to have arisen from cleavage of an incompletely blocked lysyl residue (13). However, we can now deduce that despite the

[1] Portions of this paper (including "Experimental Procedures," Figs. 1S and 2S, and Tables 1S–8S) are presented in miniprint at the end of this paper. The abbreviations used are: PTH, phenylthiohydantoin; HPLC, high performance liquid chromatography. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 83M-2934, cite the authors, and include a check or money order for $7.20 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

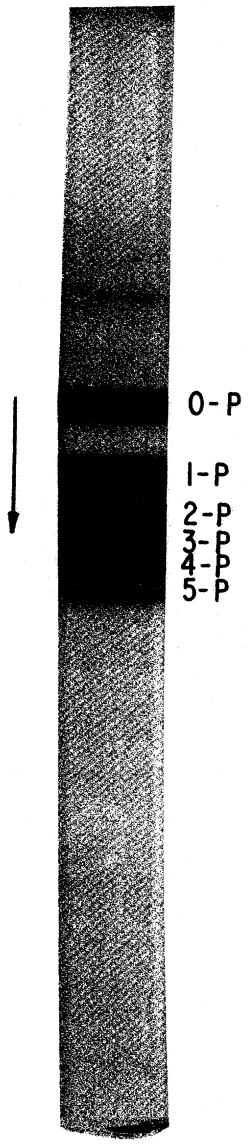

FIG. 1. **Polyacrylamide gel electrophoretic pattern (pH 9.5, 4 M urea) of human β-casein.** 0- to 5-P indicates level of phosphorylation.

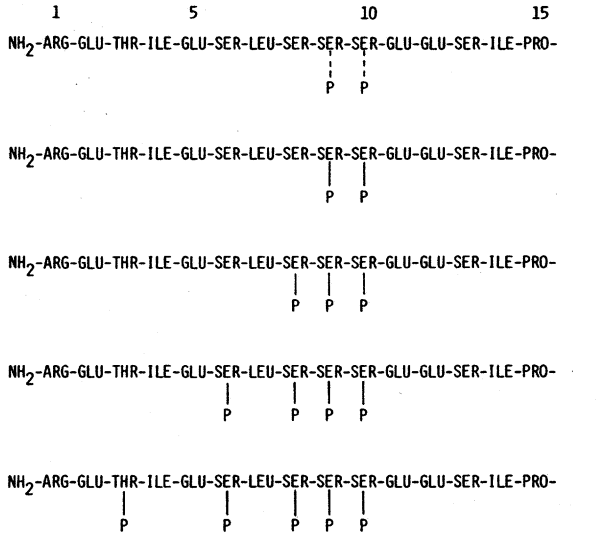

FIG. 2. **Location of phosphate groups on the 1-P to 5-P forms of human β-casein.** See text for explanation of the *dashed lines.*

use of L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated trypsin in the digest, these peptides are the result of a chymotryptic-type split between leucine 49 and isoleucine 50. When the smallest cyanogen bromide peptide, CB-2, was subjected to automated Edman analysis, the first 14 of its 16 residues were identified and the terminal arginine of T-1B thereby located. The two remaining amino acids of CB-2 by composition, valine and methionine, serve as overlaps to the beginning of the next large fragment, T-2. Considering the accurate composition of CB-2, the fact that methionine must be the carboxyl-terminal residue and that arginine 98 marks the cleavage point between T-1B and T-2, it is most unlikely that any residues have been missed here.

Elucidation of the last third of the human β-casein molecule presented a somewhat unique problem when attempts at automated Edman degradation of CB-4 failed. It was obvious that the amino terminus was blocked, the most likely culprit being pyrrolidonecarboxylic acid (16) arising from the cyclization of glutamine 136. Treatment with pyrrolidonecarboxylyl peptidase as recommended by Doolittle (17) was unsuccessful; the blocked residue persisted. During the initial purification of this cyanogen bromide fragment (13), fractions from the peak of interest were subjected to polyacrylamide gel electrophoresis at pH 4.3, 8 M urea. It was observed that the beginning of the peak consisted of one gel band, the middle a mixture of two, and the descending area the second band of increased mobility. All these fractions had identical amino acid compositions and were designated CB-4 but not pooled. Since such multiple banding of purified peptides might be ascribed to amide differences (13), it seemed possible that the two bands represented a separation of a species with intact glutamine as the amino terminus and another with L-2-pyrrolidonecarboxylic acid in that position. Confirming this speculation, sequence analysis of the two-banded fraction and that of the purified component of faster mobility proceeded without difficulty.

Studies on peptide CB-4 extended the sequence data from residues 136 to 175 including the arginyl residue at 174 which marks the cleavage point for the final tryptic fragment, T-3. The latter peptide is devoid of lysine and even in the presence of Polybrene was not well retained in the spinning cup. The poor yields caused some uncertainty near the end of the sequence and amide or acid assignments could not be made for positions 205 and 207. The carboxyl-terminal three residues were assigned by carboxypeptidase digestion of the whole molecule.

The total composition by sequence is 212 amino acids as compared to the reported value of 210 (5, 13). The changes involve one more serine (10 *versus* 9), two additional proline residues (41 *versus* 39), and one less leucine (25 *versus* 26). Except for serine, these adjustments to the earlier composition involve amino acids present in high numbers, a not unexpected finding.

Relevant to this composition, during Edman degradation position 142 (Table 7S) yielded equal amounts of isoleucine and proline in direct contrast to the preceding and following residues where only one phenylthiohydantoin derivative was evident. This suggests genetic polymorphism at this site involving no change of charge—a so called "silent" genetic variant. A similar occurrence at 169 where equal amounts of leucine and glutamine were found may be another uncharged substitution, but background accumulation at that stage of a sequence experiment prevents an unequivocal assertion. In-

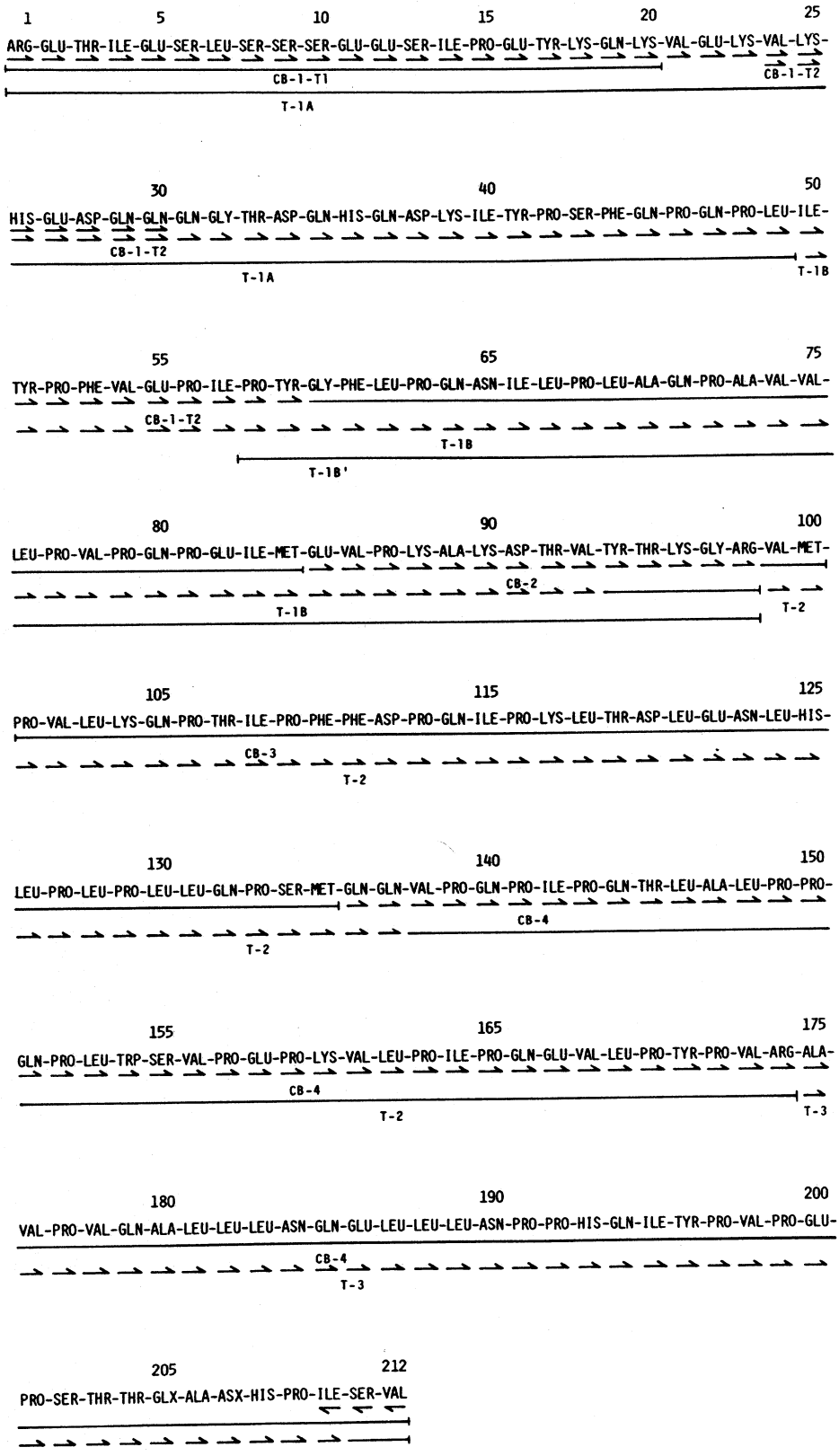

FIG. 3. **Strategy for determination of the amino acid sequence of human β-casein.** The cyanogen bromide (*CB*), trifluoroacetyl-blocked tryptic (*T*), and CB-1-T peptides are indicated below the line to which they refer. The *forward arrows* indicate those peptides sequenced; the *reverse arrows* indicate the use of carboxypeptidases. *Solid lines* show residues placed by peptide composition only.

volvement of these amino acids in such polymorphism would contribute to the uncertainty in totals discussed above.

DISCUSSION

Phosphorylation—The sequence cluster at residues 5–12 of human β-casein corresponding to the site of phosphorylation is also found at varying locations in bovine β-, α_{s1}-, and α_{s2}-caseins (11, 15). According to Mercier *et al.* (18) (see the review (15) for details), the configuration Thr/Ser-*X*-*A*, where *X* represents any amino acid, and *A*, an acidic residue, is required for casein phosphorylation. Sites where *A* is a dicarboxylic amino acid are termed primary sites; and those with phosphoserine in that position, secondary sites. From Fig. 2, note that human β-casein contains primary phosphorylation

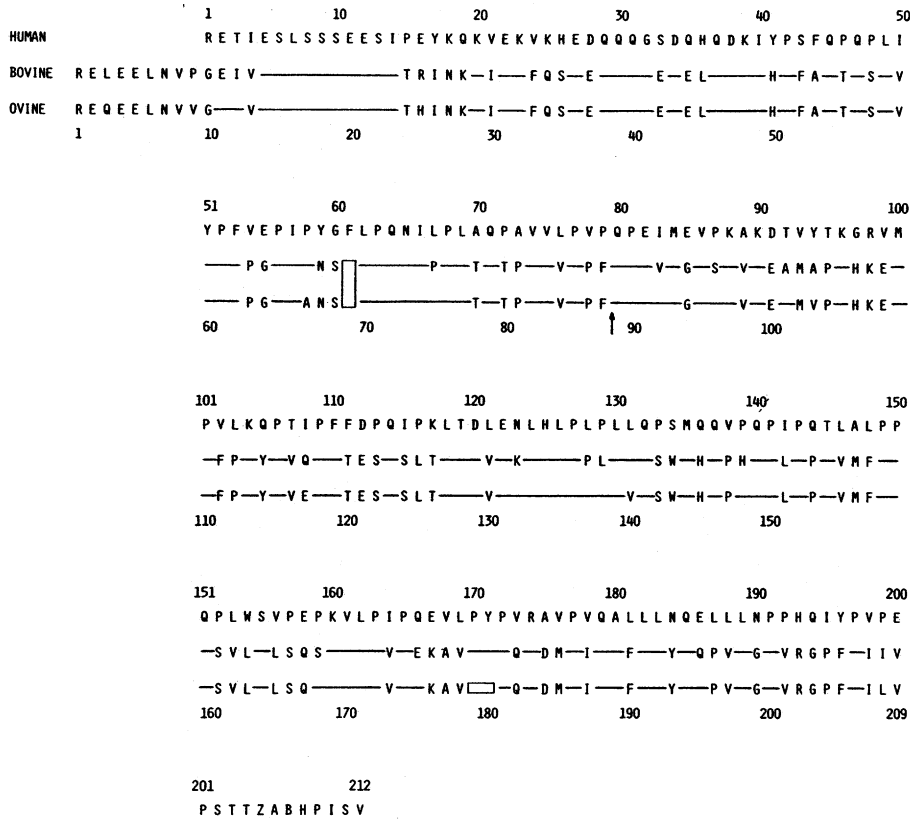

FIG. 4. **Sequences of β-caseins aligned for homologies.** *Numbering* at the *top* refers to the human sequence and at the *bottom*, the bovine homologue. Deletion in the alignment is indicated by ▭ and the presence of an extra residue by ↑. The standard one-letter symbols for amino acids are used (28). A *solid line* indicates density with the human sequence. The bovine BA[2] and ovine B[1] sequences were determined by other investigators (11, 12).

sites at positions 3, 9, and 10 and secondary sites at positions 6 and 8. As mentioned earlier, the finding of phosphorus only on specific seryl and threonyl residues in forms phosphorylated at different levels indicates that the components are indeed homogeneous with respect to phosphorylated sites and not merely charge class mixtures containing phosphate groups randomly distributed. The 1-P form alone is a mixture, with half the molelcules phosphorylated at position 9 and half at position 10. Therefore, in this case only, these two sites are equivalent; each is followed $N + 2$ by glutamic acid and the casein kinase cannot distinguish between serines 9 and 10. After this, for the 2-, 3-, 4-, and 5-P forms, phosphate addition appears to be stepwise and there are no intermediate forms. Possible phosphate migration after incorporation seems to be ruled out by its absence on the neighboring seryl residue 8 in the biphosphorylated molecule.

The reason for the existence of multiphosphorylated forms in human β-casein as compared with the fully phosphorylated cow homologue remains unclear. Possible explanations involve single or multiple kinase or phosphatase activity during biosynthesis or a variability in transport mechanism between the two species. An additional point to be considered is the appearance of quantitative variation among the six β-casein components. The pattern in Fig. 1 where the 2-P and 4-P forms show greater intensities than the others is common for many individual and pooled milk samples. Differing band intensities have been ascribed to genetic control (19). During our screening studies of individual samples, variations in band intensity patterns have been encountered. Other factors, such as stage of lactation, maternal health, or altered chemistry of calcium and phosphate, may also play a role in governing the quantitative variation. Correlation of this observation with the nutritional adequacy for the preterm infant in terms of calcium and phosphate availability may lead to the elucidation of some cause/effect relationship.

Sequence Comparison—Fig. 4 shows a comparison of the human β-casein sequence with the ruminant homologues, bovine (11) and ovine (12), the only other β-caseins for which such data are available. To align the structures it is necessary to place the amino terminus of the human molecule at position 10 of the other two sequences. A compensating carboxyl-terminal extension of the human protein provides for the approximate size equivalency, 212 *versus* 209 residues. This "frame shift" relationship is rather unusual, and considering that several residues at the amino and carboxyl termini of the three proteins are identical, the mutation may exist in the genome and could reflect a difference in the number and location of exons. There is an insert of an extra residue at human position 61 and a deletion between human 79 and 80 (bovine residue 88). In terms of amino acid identities between the homologues, the value is 47% for the human-bovine comparison, 51% for human-ovine, and 90% for the more closely related species, bovine-ovine. Most of the differences between the three proteins represent single base changes. All three molecules, characteristic of caseins in general, contain large amounts of proline (39–41 residues/mol). Over 60% of these prolyl residues are identically situated, showing conservation of the teleological requirement for openness and therefore, maximum digestibility. If the human and bovine β-casein sequences are aligned as in Fig. 4 and their secondary structures compared, the human homologue exhibits even more aperiodicity than the bovine protein (11, 20), previously reported to contain a low proportion of its residues in periodic conformation. Comparison of the hydrophobicity profiles of the molecules reveals no striking differences. Details of these comparisons will be presented elsewhere. Charge distribution along the three sequences is again quite similar—a perhaps predictable observation considering that each of these proteins is packaged by protein interaction for delivery via the "casein micelle." Several explanations are extant for the bo-

vine micelle structure (21, 22), but the dearth of knowledge of the components of the human casein system does not allow straightforward extrapolations.

Seven genetic variants of bovine β-casein have been reported (22), all demonstrable by acid or alkaline polyacrylamide gel electrophoresis. These variants, containing one or more charged residue differences, have been confirmed as truly genetic by gel pattern analysis of dam, sire, daughter relationships. For substitutions of uncharged residues, as seems to be the case in human β-casein, there is no easy way to demonstrate this phenomenon. Careful sequence analysis with good quantitation should suffice to confirm the existence of heterozygosity in the human protein. Voglino *et al.* (19), from an analysis of gel band patterns and intensities (similar to Fig. 1), propose the existence of a charged residue substitution in human β-casein but they have not isolated or characterized any material to support this contention. Since milk for the present sequence study was obtained from a single individual (5) and the multibanded pattern of Fig. 1 is due only to differences in phosphorylation, we conclude that we are dealing with heterozygous β-casein, *i.e.* genetic polymorphism involving uncharged amino acids.

Recently Monti and Jolles (23) reported on a temperature-sensitive human milk whey protein, referred to as galactothermin, which they concluded was intact 0-P β-casein. We suggest that solubility at pH 4.6 allows distribution of this form between the whey and casein fractions.

Peptides with opioid activity have been isolated from bovine caseins by several investigators (24–27). One of these, β-casomorphin (25), is equivalent to residues 60–66 of bovine β-casein, and has been prepared from an enzyme digest of whole casein. The human homologue contains several sequences characteristic of such peptides and studies are underway to test these fragments for activity.

REFERENCES

1. Committee on Nutrition, American Academy of Pediatrics (1980) *Pediatrics* **65,** 854–857
2. Ogra, P. L., and Greene, H. L. (1982) *Pediatr. Res.* **16,** 266–271
3. Lönnerdal, B., Forsum, E., and Hambraeus, L. (1976) *Nutr. Rep. Int.* **13,** 125–134
4. Hambraeus, L., Lönnerdal, B., Forsum, E., and Gebre-Medhin, M. (1978) *Acta. Pediatr. Scand.* **67,** 561–565
5. Groves, M. L., and Gordon, W. G. (1970) *Arch. Biochem. Biophys.* **140,** 47–51
6. Nagasawa, T., Kiyosawa, I., and Kuwahara, K. (1970) *J. Dairy Sci.* **53,** 136–145
7. Gross, S. J. (1983) *N. Engl. J. Med.* **308,** 237–241
8. Chessex, P., Reichman, B., Verellen, G., Putet, G., Smith, J. M., Heim, T., and Swyer, P. R. (1983) *J. Pediatr.* **102,** 107–112
9. Lemons, J. A., Moye, L., Hall, D., and Simmons, M. (1982) *Pediatr. Res.* **16,** 113–117
10. Greer, F. R., Steichen, J. J., and Tsang, R. C. (1982) *Am. J. Dis. Child.* **136,** 581–583
11. Ribadeau Dumas, B., Brignon, G., Grosclaude, F., and Mercier, J.-C. (1972) *Eur. J. Biochem.* **25,** 505–514
12. Richardson, B. C., and Mercier, J.-C. (1979) *Eur. J. Biochem.* **99,** 285–297
13. Groves, M. L., and Greenberg, R. (1978) *Biochem. J.* **169,** 337–342
14. Greenberg, R., Groves, M. L., and Peterson, R. F. (1976) *J. Dairy Sci.* **59,** 1016–1018
15. Mercier, J.-C. (1981) *Biochimie* **63,** 1–17
16. Doolittle, R. F., and Armentrout, R. W. (1968) *Biochemistry* **7,** 516–521
17. Doolittle, R. F. (1970) *Methods Enzymol.* **19,** 555–569
18. Mercier, J.-C., Grosclaude, F., and Ribadeau Dumas, B. (1971) *Eur. J. Biochem.* **23,** 41–51
19. Voglino, G. F., Caone, M., and Ponzone, A. (1975) *Biomedicine (Paris)* **23,** 344–345
20. Creamer, L. K., Richardson, T., and Parry, D. A. D. (1981) *Arch. Biochem. Biophys.* **211,** 689–696
21. Farrell, H. M., Jr. (1973) *J. Dairy Sci.* **56,** 1195–1206
22. Whitney, R. McL., Brunner, J. R., Ebner, K. E., Farrell, H. M., Jr., Josephson, R. V., Morr, C. B., and Swaisgood, H. E. (1976) *J. Dairy Sci.* **59,** 795–815
23. Monti, J. C., and Jolles, P. (1982) *Experientia (Basel)* **38,** 1211–1213
24. Brantl, V., Teschemacher, H., Henschen, A., and Lottspeich, F. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360,** 1211–1216
25. Henschen, A., Lottspeich, F., Brantl, V., and Teschemacher, H. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360,** 1217–1224
26. Zioudrou, C., Streaty, R. A., and Klee, W. A. (1979) *J. Biol. Chem.* **254,** 2446–2449
27. Brantl, V., and Teschemacher, H. (1982) *Milchwissenschaft* **37,** 581–583
28. IUPAC-IUB Commission on Biochemical Nomenclature (1968) *J. Biol. Chem.* **243,** 3557–3559
29. Davis, B. J. (1964) *Ann. N. Y. Acad. Sci.* **121,** 404–427
30. Greenberg, R., and Groves, M. L. (1979) *J. Dairy Res.* **46,** 235–239
31. Pisano, J. J., Bronzert, T. J., and Brewer, H. B., Jr. (1972) *Anal. Biochem.* **45,** 43–59
32. Jeppson, J. O., and Sjoquist, J. (1967) *Anal. Biochem.* **18,** 264–269
33. Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., and Ballantyne, D. C. (1971) *Biochemistry* **10,** 4912–4921

Supplementary Material to:

HUMAN β-CASEIN: AMINO ACID SEQUENCE AND IDENTIFICATION OF PHOSPHORYLATION SITES

Rae Greenberg, Merton L. Groves, and Harold J. Dower

EXPERIMENTAL PROCEDURES

The whole β-casein fraction from mature human milk was prepared as described previously by Groves and Gordon (5). Before chromatography, casein solutions were heated at 95 °C for 5 min to eliminate any possible proteolysis during the fractionation procedure. Separation and purification of the components were carried out on DEAE-cellulose by stepwise column chromatography in phosphate buffer (.005 to .10 M) at pH 8.3 (5). Six pure proteins having identical amino acid compositions but differing in phosphate content from zero to five residues per molecule were isolated by Groves and Gordon (5) and used for sequence studies.

Polyacrylamide Gel Electrophoresis

To monitor the purity and identity of the proteins and peptide fractions, polyacrylamide disc gel electrophoresis was carried out by the method described by Davis (29) with 7.5% gels at either pH 9.5 or pH 4.3. Occasionally, the alkaline gels were run in the presence of 4 M urea and the pH 4.3 gels, 8 M urea.

Amino Acid Analysis

Protein and peptide samples were hydrolyzed for 24 h with 5.7 N HCl containing phenol (.05%) in sealed evacuated tubes. Analysis were performed on a Beckman 119 CL amino acid analyzer and results reported as molar ratios.

Peptide Purification

Isolation and purification of the four peptides (CB-1 through CB-4) resulting from the cyanogen bromide cleavage of human β-casein were accomplished by Bio-Gel P30 and DEAE-cellulose chromatography at pH 8.3 in NH_4HCO_3 (13). Tryptic cleavage of trifluoracetylated human β-casein produced peptides T-1A, T-1B, T-2, and T-3 (13). The compositions of these two sets of peptides are shown in Table 1S. Peptide T-1B', also isolated and purified, is slightly smaller than T-1B and represents a nonspecific cleavage product.

Digestion of CB-1 with trypsin for 45 min at an enzyme-peptide ratio of 1:70 released two fragments which were separated and purified on Sephadex G25 in 0.1 M acetic acid. Compositions of these peptides designated CB-1-T1 and CB-1-T2 are included in Table 1S. Fig. 1S presents the arrangement of these peptides to comprise the molecule.

Carboxypeptidase Digestion

Carboxypeptidases A and B (Worthington) were used at ratio of 1:50 at pH 8.2 unbuffered systems. Digestion of the whole molecule under these conditions liberated only three residues.

Amino Acid Sequence

Sequencing was carried out on a Beckman 890 C sequencer using the double cleavage 1 M Quadrol protein program of 042772 for examination of the intact molecule and studies involving the identification and quantitation of phosphorylated residues (30). To accomplish the latter, either the entire sequencer output at each seryl or threonyl residue, or the aqueous layer from the PTH conversion step was subjected to microphosphorous analysis. For the majority of the large peptide sequencing, a .25 M Quadrol single cleavage peptide/protein program using polybrene and a simultaneous benzene and ethyl acetate wash served well. At the beginning of this study identification of PTH amino acids was conducted by gas and thin layer chromatography (31, 32) and/or hydrolysis with HI back to the parent amino acid (33).

For identification and quantitation during the peptide sequencing experiments, an HPLC method developed by us permitted resolution of 19 PTH amino acids in 15 min with an injection turn around time of 20 min (Fig. 2S). The presence of serine was still confirmed by HI hydrolysis and amino acid analysis.

All protein and peptide samples were sequenced at least twice and more often three or four times. With sufficient material available for sequencing (200 to 500 n moles per run) and yields ranging from 93% to 96%, there was no difficulty obtaining meaningful information up to 40 cycles. Results needed to deduce the sequence of human β-casein are shown in Tables 2S to 8S.

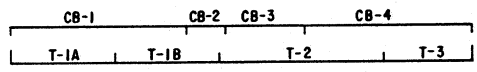

Fig. 1S. Arrangement of cyanogen bromide (CB) and Tryptic (T) peptides.

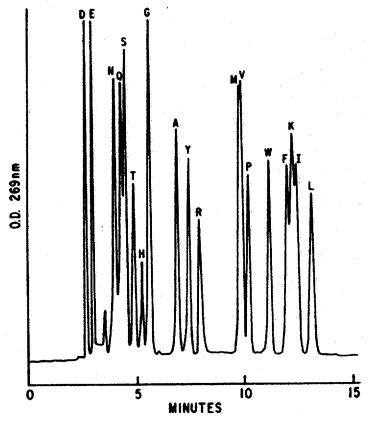

Fig. 2S. HPLC separation of PTH-amino acids on a Brownlee RP-18 column (5 μ, 4 X 250 mm). Peaks are labelled with the one letter code normally used to denote the parent amino acids (38). The elution conditions were as follows: Solvent A, 0.015 M sodium acetate, pH 4.5; solvent B, acetonitrite; linear gradient from 36% B to 44% B over 10 min at a flow rate of 1.0 ml/min for the first 5 min and 1.5 ml/min for the remainder. Temperature—35° C.

Table 1S

Amino acid composition of cyanogen bromide (CB) and tryptic (T) peptides

	CB-1	CB-2	CB-3	CB-4	T-1A	T-1B	T-2	T-3	CB-1-T1	CB-1-T2
Asp	4.0(4)[a]	1.2(1)	3.1(3)	3.0(3)	2.8(3)	2.4(2)	3.2(3)	3.0(3)		3.8(4)
Thr	2.0(2)	2.1(2)	1.9(2)	2.8(3)	1.7(2)	2.0(2)	2.9(3)	1.9(2)	0.7(1)	0.8(1)
Ser	5.0(6)		1.1(1)	1.8(3)	4.5(6)	1.1(0)	2.2(2)	1.4(2)	4.6(5)	1.1(1)
Glu	19.4(20)	1.5(1)	4.4(4)	13.7(14)	13.3(15)	6.5(6)	12.1(12)	6.0(6)	6.2(6)	13.7(13)
Pro	10.3(13)	1.2(1)	8.3(8)	17.1(19)	4.3(4)	9.2(10)	19.0(20)	7.1(7)	1.1(1)	12.6(12)
Gly	2.1(2)	1.1(1)			1.4(1)	2.9(2)	0.6(0)	0.7(0)		2.0(2)
Ala	2.1(2)	1.1(1)		3.9(4)		3.1(3)	1.3(1)	3.0(3)		2.0(2)
Val	5.3(6)	3.3(3)	1.6(1)	8.9(9)	2.3(2)	5.3(6)	6.5(7)	3.7(4)		4.7(4)
Met	1	1	1			0.7(1)	1.7(2)			
Ile	5.8(7)		2.0(2)	3.7(4)	3.1(3)	3.5(4)	3.9(4)		2.1(2)	2.1(2)	4.6(5)
Leu	5.3(6)		9.0(8)	10.9(11)	2.3(2)	3.6(4)	12.8(13)	7.2(6)	1.1(1)	4.9(5)
Tyr	3.8(4)	1.0(1)		1.9(2)	2.2(2)	2.5(3)	1.1(1)	1.0(1)	1.0(1)	2.7(3)
Phe	2.6(3)		1.7(2)		1.1(1)	1.7(2)	1.9(2)			2.7(3)
Lys	4.9(5)	3.1(3)	2.0(2)	1.1(1)	4.0(5)	3.0(3)	3.0(3)	0.5(0)	1.8(2)	2.0(2)
His	1.8(2)		1.0(1)	1.8(2)	1.6(2)		0.9(1)	1.8(2)		2.0(2)
Arg	1.0(1)	1.0(1)		1.0(1)	0.7(1)	0.9(1)	0.9(1)			
Try[b]				1			1			

[a] Values in parenthesis from sequence data.

[b] Qualitative determination only.

Table 2S

Amino Acid Sequence of Intact Human β-casein

Cycle	Amino acid	Yield n moles	Method of detection[1]
1	Arg		e
2	Glu		a,b,d
3	Thr		a,b,c,d
4	Ile	214	a,d
5	Glu		a,b
6	Ser		a,c
7	Leu		a,b
8	Ser		a,c
9	Ser		a,c
10	Ser		a,c
11	Glu		a,b
12	Glu		a,b
13	Ser		a,c
14	Ile	113	a,b
15	Pro		a
16	Glu		a,b
17	Tyr		a,b
18	Lys		a,b
19	Gln		a,b,d,
20	Lys		a,b,d
21	Val	96	a,b
22	Glu		a,b
23	Lys		a,b
24	Val	87	a,b
25	Lys		a,b
26	His		d,e
27	Glu		a,b
28	Asp		b,c,d
29	Gln		b
30	Gln		b
31	Gln		b
32	Gly		b

300 n moles applied; yield Ile 4 → Val 24, 95.5%

[1]
a - Gas chromatography

b - Thin layer chromatography

c - Amino acid analysis after HI hydrolysis

d - HPLC

e - Spot tests - histidine - diazotized sulfanilic acid; arginine - phenanthrene quinone

Table 3S

Amino Acid Sequence of CB-1-T2

Cycle	Amino acid	Yield n moles	Method of detection
1	Val	300	a,d
2	Lys	269	a,d
3	His		d,e
4	Glu		a
5	Asp		a,c,d
6	Gln		a,d
7	Gln		a,d
8	Gln		a,d
9	Gly	260	a
10	Thr		c,d
11	Asp		a,c,d
12	Gln		a,d
13	His		d,e
14	Gln		d
15	Asp		d
16	Lys		d
17	Ile		a,d
18	Tyr	100	a,d
19	Pro	83	a,d,
20	Ser		c,d
21	Phe	81	a,d
22	Gln		d
23	Pro	87	a,d
24	Gln		d
25	Pro	93	a,d
26	Leu	47	a,d
27	Ile	39	a,d
28	Tyr	41	a,d
29	Pro		a,d
30	Phe	32	a,d
31	Val	28	a,d
32	Glu		d
33	Pro	28	a,d
34	Ile	20	a,d
35	Pro		d
36	Tyr		d

416 n moles applied; initial yield 72%, repetitive yield Val 1 → 31, 93%

See footnote - Table 2S

Table 4S

Amino Acid Sequence of T-1B

Cycle	Amino acid	Yield n moles	Method of detection
1	Ile	340	a
2	Tyr		a
3	Pro		a
4	Phe	268	a,d
5	Val	259	a,d
6	Glu		a,d
7	Pro		a,d
8	Ile		a
9	Pro		a
10	Tyr	260	a,d
11	Gly		d
12	Phe	219	d
13	Leu		a
14	Pro		a
15	Gln		a,d
16	Asn		a,d
17	Ile		a
18	Leu		a
19	Pro	139	a,d
20	Leu	129	d
21	Ala	121	d
22	Gln		d
23	Pro	101	d
24	Ala	99	d
25	Val	99	a,d
26	Val		a,d
27	Leu	70	a,d
28	Pro	65	a,d
29	Val	74	a,d
30	Pro	83	a,d
31	Gln		d
32	Pro		a,d
33	Glu		a,d
34	Ile	42	a,d
35	Met	39	a,d
36	Glu		d
37	Val	37	a,d
38	Pro Val	27 24	a,d
39	Lys		d
40	Ala	26	a,d
41	Lys		d
42	Asp		d
43	Thr		d
44	Val	17	d

500 n moles applied; initial yield 68%, repetitive yield Val 5 → 44, 93.2%

See footnote - Table 2S

Table 5S

Amino Acid Sequence of CB-2

Cycle	Amino acid	Yields n moles	Method of detection
1	Glu	483	d
2	Val		a
3	Pro		a
4	Lys	213	d
5	Ala		a
6	Lys	160	d
7	Asp		d
8	Thr		d
9	Val		a
10	Tyr	97	a,d
11	Thr		d
12	Lys	48	d
13	Gly	12	d
14	Arg		e

1000 n moles applied; initial yield 48%; repetitive yield Glu 1 → Gly 13, 73%

See footnote—Table 2S

Table 6S

Amino Acid Sequence of T-2

Cycle	Amino acid	Yield n moles	Method of detection
1	Val		a,d
2	Met	328	a,d
3	Pro	267	a,d
4	Val	235	a,d
5	Leu		a,d
6	Lys	191	d
7	Gln		d
8	Pro	204	d
9	Thr		d
10	Ile		a,d
11	Pro	173	d
12	Phe	146	a,d
13	Phe	145	a,d
14	Asp	107	a,d
15	Pro		a,d
16	Gln		d
17	Ile		a,d
18	Pro	77	a,d
19	Lys		d
20	Leu	85	d
21	Thr		d
22	Asp	62	d
23	Leu	62	d
24	Glu	60	d
25	Asn		d
26	Leu	44	d
27	His		d,e
28	Leu	48	d
29	Pro	27	a,d
30	Leu	29	a,d
31	Pro	25	a,d
32	Leu	27	a,d
33	Leu	24	a,d
34	Gln		d
35	Pro	20	a,d
36	Ser		c,d
37	Met	10	d
38	Gln		d
39	Gln		d
40	Val	9	d

500 n moles applied; repetitive yield Met 2 → 37, 91%

See footnote - Table 2S

Table 7S

Amino Acid Sequence of CB-4

Cycle	Amino acid	Yield n moles	Method of detection
1	Gln	154	d
2	Gln	138	d
3	Val	130	d
4	Pro	127	d
5	Gln	78	d
6	Pro	108	d
7	Ile Pro	42 43	d
8	Pro	87	d
9	Gln		d
10	Thr		d
11	Leu	35	d
12	Ala	25	d
13	Leu	39	d
14	Pro	54	d
15	Pro	59	d
16	Gln		d
17	Pro	43	d
18	Leu	22	d
19	Trp		d
20	Ser		c,d
21	Val	20	d
22	Pro	31	d
23	Glu	16	d
24	Pro	34	d
25	Lys		d
26	Val	17	d
27	Leu	9	d
28	Pro	25	d
29	Ile		d
30	Pro	18	d
31	Gln		d
32	Glu		d
33	Val		d
34	Leu Gln		d
35	Pro	13	d
36	Tyr		d
37	Pro		d
38	Val		d
39	Arg		d,e
40	Ala		d

Yield Leu 11 → 18, 94%

See footnote - Table 2S

Table 8S

Amino Acid Sequence of T-3

Cycle	Amino acid	Yield n moles	Method of detection
1	Ala	495	a,d
2	Val	487	a,d
3	Pro	327	a,d
4	Val	277	a,d
5	Gln		c,d
6	Ala		a,d
7	Leu		a,d
8	Leu	142	a,d
9	Leu	126	d
10	Asn		d
11	Gln		c,d
12	Glu		d
13	Leu	65	d
14	Leu	72	d
15	Leu	77	d
16	Asn		d
17	Pro	37	d
18	Pro	31	d
19	His		d,e
20	Gln		d
21	Ile	21	d
22	Tyr	12	d
23	Pro	12	d
24	Val	10	d
25	Pro		d
26	Glu		d
27	Pro		d
28	Ser		c,d
29	Thr		d
30	Thr		d
31	Glx		d
32	Ala		d
33	Asx		d
34	His		d,e
35	Pro		d
36	Ile		d

660 n moles applied; repetitive yield Val 4 → 24, 85%

See footnote - Table 2S